

Robust Face Recognition via Sparse Representation

Allen Y. Yang
<yang@eecs.berkeley.edu>

April 18, 2008, NIST



Face Recognition: “Where amazing happens”

Face Recognition: “Where amazing happens”



Figure: Steve Nash, Kevin Garnett, Jason Kidd.

Sparse Representation

Sparsity

A signal is sparse if most of its coefficients are (approximately) zero.



Sparse Representation

Sparsity

A signal is sparse if most of its coefficients are (approximately) zero.

① Sparsity in frequency domain



Figure: 2-D DCT transform.

② Sparsity in spatial domain

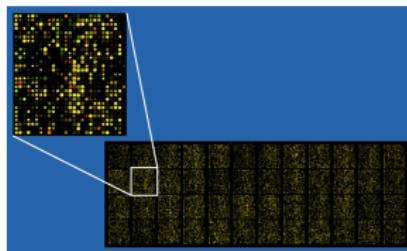
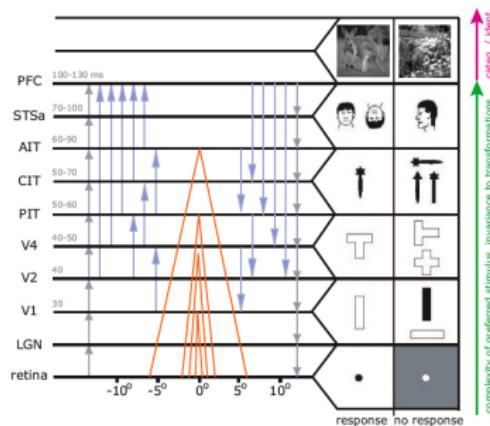
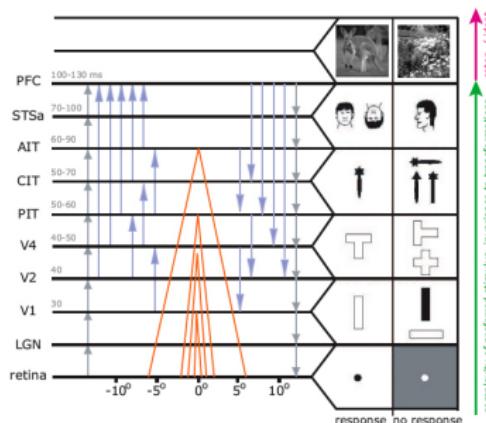


Figure: Gene microarray data.

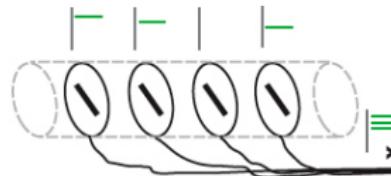
- Sparsity in human visual cortex [Olshausen & Field 1997, Serre & Poggio 2006]



- Sparsity in human visual cortex [Olshausen & Field 1997, Serre & Poggio 2006]



- ① **Feed-forward:** No iterative feedback loop.
- ② **Redundancy:** Average 80-200 neurons for each feature representation.
- ③ **Recognition:** Information exchange between stages is not about individual neurons, but rather **how many neurons as a group fire together**.



Problem Formulation

① Notation

- Training: For K classes, collect training samples $\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,n_1}\}, \dots, \{\mathbf{v}_{K,1}, \dots, \mathbf{v}_{K,n_K}\} \in \mathbb{R}^D$.
- Test: Present a new $\mathbf{y} \in \mathbb{R}^D$, solve for $\text{label}(\mathbf{y}) \in [1, 2, \dots, K]$.

Problem Formulation

① Notation

- Training: For K classes, collect training samples $\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,n_1}\}, \dots, \{\mathbf{v}_{K,1}, \dots, \mathbf{v}_{K,n_K}\} \in \mathbb{R}^D$.
- Test: Present a new $\mathbf{y} \in \mathbb{R}^D$, solve for label(\mathbf{y}) $\in [1, 2, \dots, K]$.

② Data representation in (long) vector form via stacking

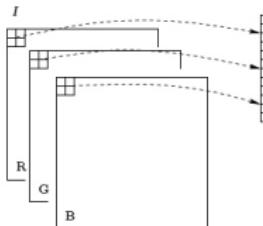


Figure: Assume 3-channel 640×480 image, $D = 3 \cdot 640 \cdot 480$.

Problem Formulation

① Notation

- Training: For K classes, collect training samples $\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,n_1}\}, \dots, \{\mathbf{v}_{K,1}, \dots, \mathbf{v}_{K,n_K}\} \in \mathbb{R}^D$.
- Test: Present a new $\mathbf{y} \in \mathbb{R}^D$, solve for label(\mathbf{y}) $\in [1, 2, \dots, K]$.

② Data representation in (long) vector form via stacking

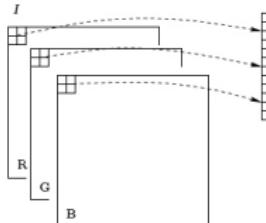
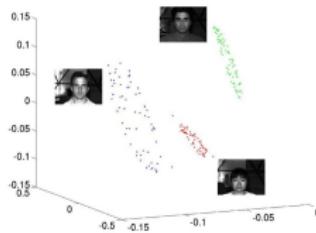


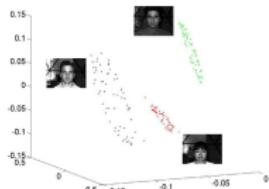
Figure: Assume 3-channel 640×480 image, $D = 3 \cdot 640 \cdot 480$.

③ Mixture subspace model for face recognition [Belhumeur et al. 1997, Basri & Jacobs 2003]



Classification of Mixture Subspace Model

① Assume \mathbf{y} belongs to Class i

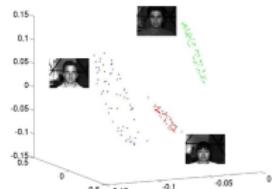


$$\begin{aligned}\mathbf{y} &= \alpha_{i,1}\mathbf{v}_{i,1} + \alpha_{i,2}\mathbf{v}_{i,2} + \cdots + \alpha_{i,n_i}\mathbf{v}_{i,n_i}, \\ &= \mathbf{A}_i\boldsymbol{\alpha}_i,\end{aligned}$$

where $\mathbf{A}_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \dots, \mathbf{v}_{i,n_i}]$.

Classification of Mixture Subspace Model

- ① Assume \mathbf{y} belongs to Class i



$$\begin{aligned}\mathbf{y} &= \alpha_{i,1}\mathbf{v}_{i,1} + \alpha_{i,2}\mathbf{v}_{i,2} + \cdots + \alpha_{i,n_i}\mathbf{v}_{i,n_i}, \\ &= \mathbf{A}_i\boldsymbol{\alpha}_i,\end{aligned}$$

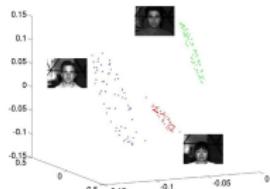
where $\mathbf{A}_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \dots, \mathbf{v}_{i,n_i}]$.

- ② Nevertheless, Class i is the **unknown** variable we need to solve:

Sparse representation $\mathbf{y} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} = \mathbf{A}\boldsymbol{\alpha} \in \mathbb{R}^{3 \cdot 640 \cdot 480}$.

Classification of Mixture Subspace Model

- ① Assume \mathbf{y} belongs to Class i



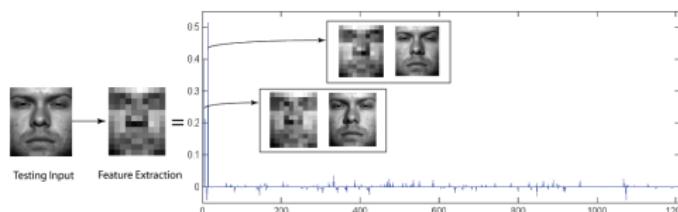
$$\begin{aligned}\mathbf{y} &= \alpha_{i,1}\mathbf{v}_{i,1} + \alpha_{i,2}\mathbf{v}_{i,2} + \cdots + \alpha_{i,n_i}\mathbf{v}_{i,n_i}, \\ &= A_i\alpha_i,\end{aligned}$$

where $A_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \dots, \mathbf{v}_{i,n_i}]$.

- ② Nevertheless, Class i is the **unknown** variable we need to solve:

Sparse representation $\mathbf{y} = [A_1, A_2, \dots, A_K] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} = A\mathbf{x} \in \mathbb{R}^{3 \cdot 640 \cdot 480}$.

- ③ $\mathbf{x}_0 = [0 \dots 0 \ \alpha_i^T \ 0 \dots 0]^T \in \mathbb{R}^n$.



Sparse representation encodes membership!

Dimensionality Reduction

- ① Construct linear projection $R \in \mathbb{R}^{d \times D}$, d is the **feature dimension**.

$$\tilde{\mathbf{y}} \doteq R\mathbf{y} = RA\mathbf{x}_0 = \tilde{A}\mathbf{x}_0 \in \mathbb{R}^d.$$

$\tilde{A} \in \mathbb{R}^{d \times n}$, but \mathbf{x}_0 is unchanged.

Dimensionality Reduction

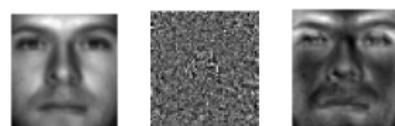
- ① Construct linear projection $R \in \mathbb{R}^{d \times D}$, d is the **feature dimension**.

$$\tilde{\mathbf{y}} \doteq R\mathbf{y} = RA\mathbf{x}_0 = \tilde{A}\mathbf{x}_0 \in \mathbb{R}^d.$$

$\tilde{A} \in \mathbb{R}^{d \times n}$, but \mathbf{x}_0 is unchanged.

- ② Holistic features

- Eigenfaces [Turk 1991]
- Fisherfaces [Belhumeur 1997]
- Laplacianfaces [He 2005]

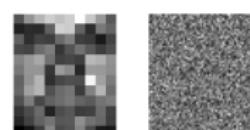


- ③ Partial features



- ④ Unconventional features

- Downsampled faces
- Random projections



ℓ^1 -Minimization

① Ideal solution: ℓ^0 -Minimization

$$(P_0) \quad \mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x}.$$

$\|\cdot\|_0$ simply counts the number of nonzero terms.
However, generally ℓ^0 -minimization is *NP-hard*.

ℓ^1 -Minimization

- ① Ideal solution: ℓ^0 -Minimization

$$(P_0) \quad \mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x}.$$

$\|\cdot\|_0$ simply counts the number of nonzero terms.

However, generally ℓ^0 -minimization is *NP-hard*.

- ② Compressed sensing: Under mild condition, ℓ^0 -minimization is equivalent to

$$(P_1) \quad \mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x},$$

where $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_n|$.

ℓ^1 -Minimization

① Ideal solution: ℓ^0 -Minimization

$$(P_0) \quad \mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x}.$$

$\|\cdot\|_0$ simply counts the number of nonzero terms.

However, generally ℓ^0 -minimization is *NP-hard*.

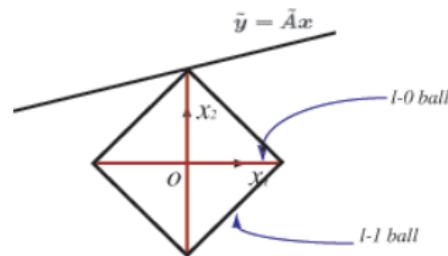
② Compressed sensing: Under mild condition, ℓ^0 -minimization is equivalent to

$$(P_1) \quad \mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x},$$

where $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_n|$.

③ ℓ^1 -Ball

- ℓ^1 -Minimization is convex.
- Solution equal to ℓ^0 -minimization.



ℓ^1 -Minimization Routines

• Matching pursuit [Mallat 1993]

- ① Find most correlated vector \mathbf{v}_i in A with \mathbf{y} : $i = \arg \max \langle \mathbf{y}, \mathbf{v}_j \rangle$.
- ② $A \leftarrow A^{(i)}$, $x_i \leftarrow \langle \mathbf{y}, \mathbf{v}_i \rangle$, $\mathbf{y} \leftarrow \mathbf{y} - x_i \mathbf{v}_i$.
- ③ Repeat until $\|\mathbf{y}\| < \epsilon$.

• Basis pursuit [Chen 1998]

- ① Start with number of sparse coefficients $m = 1$.
- ② Select m linearly independent vectors B_m in A as a basis

$$\mathbf{x}_m = B_m^\dagger \mathbf{y}.$$

- ③ Repeat swapping one basis vector in B_m with another vector not in B_m if improve $\|\mathbf{y} - B_m \mathbf{x}_m\|$.
- ④ If $\|\mathbf{y} - B_m \mathbf{x}_m\|_2 < \epsilon$, stop; Otherwise, $m \leftarrow m + 1$, repeat Step 2.

• Quadratic solvers: $\mathbf{y} = A\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^d$, where $\|\mathbf{z}\|_2 < \epsilon$

$$\mathbf{x}^* = \arg \min \{\|\mathbf{x}\|_1 + \lambda \|\mathbf{y} - A\mathbf{x}\|_2\}$$

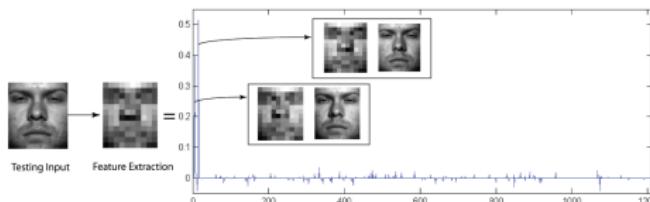
[LASSO, Second-order cone programming]: Much more expensive.

Matlab Toolboxes for ℓ^1 -Minimization

- **ℓ^1 -Magic** by Candes
- **SparseLab** by Donoho
- **cvx** by Boyd

Sparse Representation Classification

Solve $(P_1) \Rightarrow \mathbf{x}_1$.

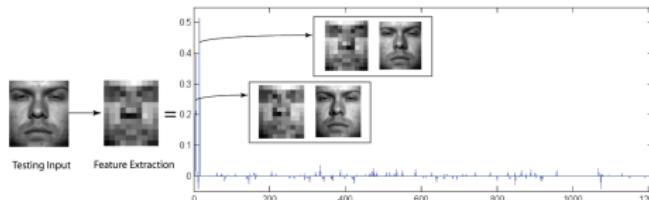


① Project \mathbf{x}_1 onto face subspaces:

$$\delta_1(\mathbf{x}_1) = \begin{bmatrix} \alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \delta_2(\mathbf{x}_1) = \begin{bmatrix} 0 \\ \alpha_2 \\ \vdots \\ 0 \end{bmatrix}, \dots, \delta_K(\mathbf{x}_1) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \alpha_K \end{bmatrix}. \quad (1)$$

Sparse Representation Classification

Solve $(P_1) \Rightarrow \mathbf{x}_1$.

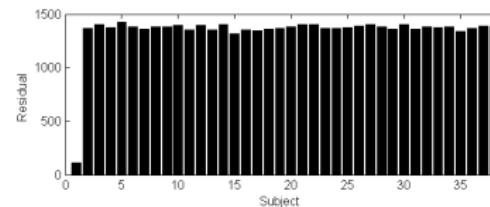


① Project \mathbf{x}_1 onto face subspaces:

$$\delta_1(\mathbf{x}_1) = \begin{bmatrix} \alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \delta_2(\mathbf{x}_1) = \begin{bmatrix} 0 \\ \alpha_2 \\ \vdots \\ 0 \end{bmatrix}, \dots, \delta_K(\mathbf{x}_1) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \alpha_K \end{bmatrix}. \quad (1)$$

② Define residual $r_i = \|\tilde{\mathbf{y}} - \tilde{A}\delta_i(\mathbf{x}_1)\|_2$ for Subject i :

- $\text{id}(\mathbf{y}) = \arg \min_{i=1, \dots, K} \{r_i\}$



Partial Features on Extended Yale B Database

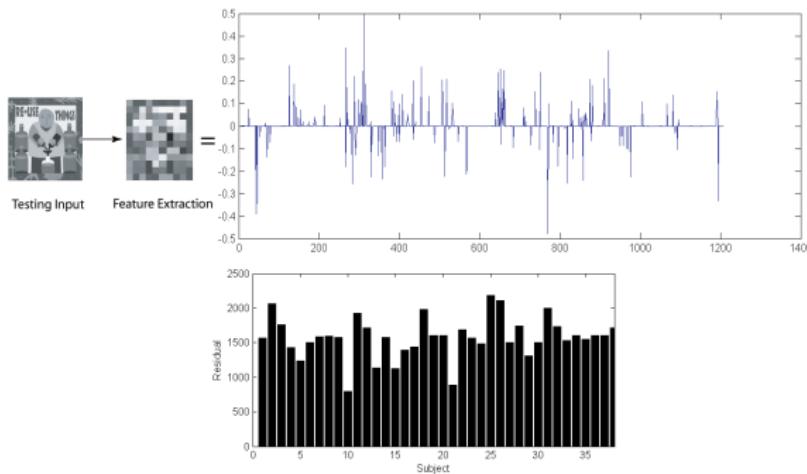


Features	Nose	Right Eye	Mouth & Chin
Dimension	4,270	5,040	12,936
SRC [%]	87.3	93.7	98.3
nearest-neighbor [%]	49.2	68.8	72.7
nearest-subspace [%]	83.7	78.6	94.4
Linear SVM [%]	70.8	85.8	95.3

SRC: sparse-representation classifier

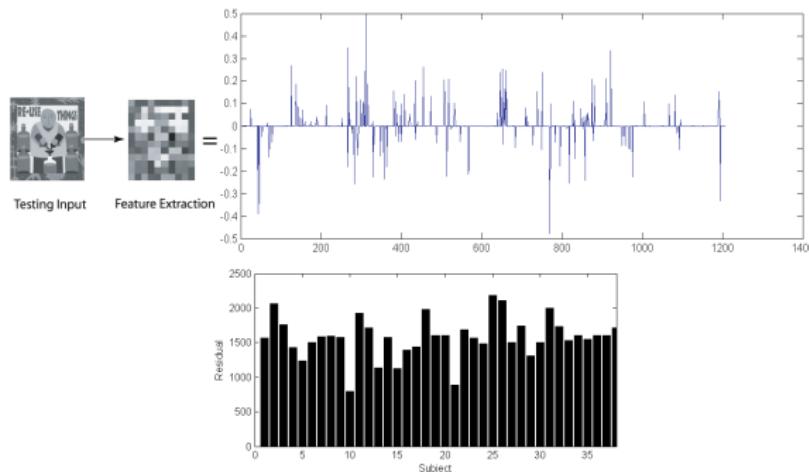
Extension I: Outlier Rejection

- ℓ^1 -Coefficients for invalid images



Extension I: Outlier Rejection

- ℓ^1 -Coefficients for invalid images



Outlier Rejection

When ℓ^1 -solution is not sparse or concentrated to one subspace, the test sample is invalid.

$$\text{Sparsity Concentration Index: } \text{SCI}(\mathbf{x}) \doteq \frac{K \cdot \max_i \|\delta_i(\mathbf{x})\|_1 / \|\mathbf{x}\|_1 - 1}{K - 1} \in [0, 1].$$

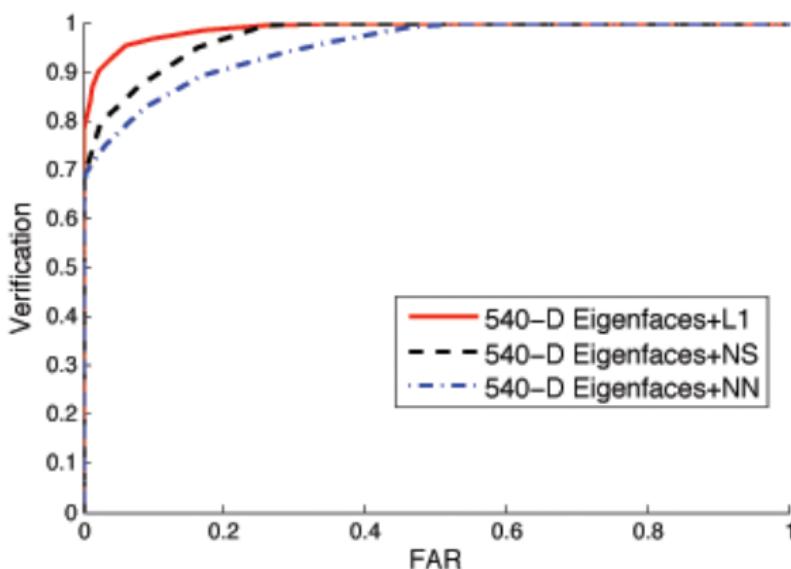
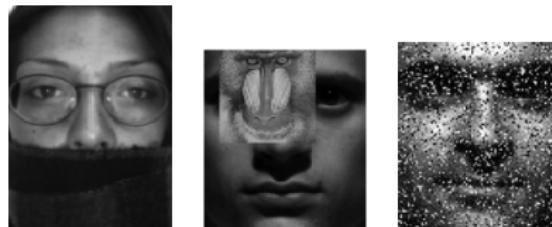


Figure: ROC curve on Eigenfaces and AR database.

Extension II: Occlusion Compensation

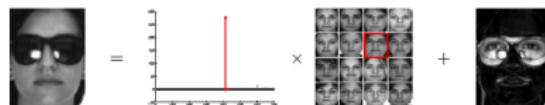


Extension II: Occlusion Compensation



① Sparse representation + sparse error

$$\mathbf{y} = \mathbf{Ax} + \mathbf{e}$$



② Occlusion compensation

$$\mathbf{y} = (\mathbf{A} \quad | \quad \mathbf{I}) \begin{pmatrix} \mathbf{x} \\ \mathbf{e} \end{pmatrix} = \mathbf{Bw}$$

AR Database



Figure: Training samples for Subject 1.

AR Database



Figure: Training samples for Subject 1.



(a) random corruption

(b) occlusion

AR Database

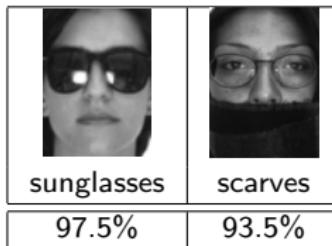


Figure: Training samples for Subject 1.



(a) random corruption

(b) occlusion



Future Directions

Open problems:

- ① Pose variation
- ② Scalability to > 1000 subjects

Future Directions

Open problems:

- ① Pose variation
- ② Scalability to > 1000 subjects

Other databases:

- ① Multi-PIE (about 350 subjects)
- ② Chinese CASPEAL (about 1000-3000 subjects)

Future Directions

Open problems:

- ① Pose variation
- ② Scalability to > 1000 subjects

Other databases:

- ① Multi-PIE (about 350 subjects)
- ② Chinese CASPEAL (about 1000-3000 subjects)

Wish list: Because few algorithm succeed under all-weather conditions (illumination, expression, pose, disguise), we are looking forward to a comprehensive database

- ① large number of subjects
- ② carefully controlled subclasses

Acknowledgments

Collaborators

- **Berkeley:** Prof. Shankar Sastry
- **UIUC:** Prof. Yi Ma, John Wright, Arvind Ganesh

Funding Support

- ARO MURI: Heterogeneous Sensor Networks (HSNs)

References

- Robust Face Recognition via Sparse Representation, (in press) **PAMI**, 2008.
- <http://www.eecs.berkeley.edu/~yang>